Statistical Inference for Atmospheric Transport Model Using Process Convolutions

Weining Zhou* and Bruno Sansó*
[zhouwn,bruno]@ams.ucsc.edu

Abstract

A computer simulator for atmospheric concetrations of chemical species, or chemical transport model, is used to simulate global ozone concentrations. Two different wind forcings are considered, one a combination of a numerical weather prediction model and observational data, the other results from a climate model. The goal is to study the impact of meteorological variability on ozone. The statistical approach that we consider consists on learning the spatial structure of ozone concentrations by using process convolutions. Then we use several Bayesian model comparison methods to determine if the two simulations can be considered as realizations of the same random field. The methods provide a quantification of the differences for each of the computer model grid cells.

Keywords: computer simulator, general circulation model, process convolutions, dimension reduction, predictive approach

1 Introduction

Ozone can be produced photochemically from nitrogen oxide and reactive organic compounds in the lower atmosphere, troposphere and stratosphere. Chemically produced ozone is a notable greenhouse gas, contributing to the extra warming of the earth. It is also a hazardous pollutant to human health, which can cause eye irritation, respiratory problems, and lung and heart diseases.

The scientists at the National Center for Atmospheric Research (NCAR) use the off-line Model for OZone And Related chemical Tracers, version2/Model of Atmospheric Transport and Chemistry (MOZART-2/MATCH) to study ozone global concentrations and transport. MOZART-2/MATCH is a global chemical transport model developed at NCAR, the Geophysical Fluid Dynamics Laboratory at Princeton University, and the Max Planck Institute of Meteorology to simulate the distribution of tropospheric ozone, its precursors and other chemical tracers. MOZART's output is driven by wind fields obtained from general circulation models (GCM) or reanalysis data sets. A general circulation model uses the equations of motion to numerically simulate changes in the global phenomena of the atmosphere, variabilities in temperature, water vapor, photolysis rate, etc. (Kasagara and Washington, 1967). Reanalysis data correspond to a combination of climate model output and observations. The National Center for Environmental Prediction (NCEP) provides a reanalysis data set (Kalnay et al., 1996) widely used by climatologists. The simulations considered in this paper were performed by Dr P. Hess, J.F. Lamarque and Dr N. Mahowald,

*Department of Applied Mathematics and Statistics, University of California, 1156 High Street, Mail Stop SOE2, Santa Cruz, CA-95064, USA.

atmospheric scientists at NCAR, using both forcings. The goal is to examine if the variability observed in ozone concentrations obtained from CAM forcings is comparable to the one produced by reanalysis winds.

While reanalysis winds represent our best understanding of the state of the atmosphere, general circulation models must be used to predict atmospheric state in past climates or future climates. These models are also used to predict the chemical state of the atmosphere in other climates. Thus here our goal is to determine whether the meteorological variability forcing the MOZART model produces similar variability in chemical constitutent distributions in the current climate whether we use our reanalysis datasets or general circulaiton model output. If so, we can use our general circulation model for future predictions of ozone (for example) with more confidence in the results.

From a statistical perspective the problem is to decide whether two sets of MOZART output correspond to realizations of the same stochastic process. Inference is complicated by the large dimensionality of the problem. In fact the output from MOZART consists of a profile for each point in a grid of 64 latitudes and 128 longitudes. That is 8,192 locations for each of 19 altitudes. A Gaussian random field, defined by a mean and a covariance matrix, can be used to model the spatial structure of the simulated ozone. However the huge amount of data and very large spatial scales raise challenging statistical analysis problems. It is impractical to perform computations or even store a gigantic covariance matrix for 8,192 locations. This would require around 8 Gb of storage.

Instead of more traditional spatial modeling methods based on the second order properties of the process, we use process convolutions (see, Higdon, 2002) to model a Gaussian process for both CAM-based and NCEP-based simulation data sets. We proceed by using a Bayesian method. This provides a whole distribution for the estimated parameters and the predicted values, fully accounting for the uncertainty in parameter estimation. We fit the model for each of the datasets separately, obtaining two sets of posterior and posterior predictive distributions. We compute the amount of overlapping of predictive intervals and Bhattacharyya distances to explore the proximity of the models. The idea is that, if the two data sets did in fact arise from the same model, then the predictive distributions should be very close. We also adopt a model comparison approach based on considering two models: one where one set of parameters describe the likelihood of all observations and one where two sets of parameters are used. The model comparison criteria include posterior predictive loss and Deviance Information Criterion (DIC).

The paper is organized as follows: the next section provides an introduction to our statistical model. Section 4 describes the model comparison methods. In the last section we present some comments and discussion.

2 Process convolutions

The process convolution approach was first brought up as a spatial moving average process (Thiébaux and Pedder, 1987; Barry and Ver Hoef, 1996). It provides a flexible representation of Gaussian processes by convolving a white noise process with an appropriate kernel. A process convolution, as defined in Higdon (2002), is given by

$$z(s) = \int_S k_\psi(u - s) dW(u) \tag{1}$$

for $s \in S$, S a subset of an Euclidean space, W a Wiener process and $k_\psi(\cdot)$ a kernel, possibly depending on a low dimensional parameter ψ. The covariance of the process $z(s)$ defined in Equation

(1) is given by the convolution of the kernel with itself, so

$$c(s, s') = \mathrm{cov}(z(s), z(s')) = \int_S k_\psi(u - d) k_\psi(u) du \ ,$$

where $d = s - s'$. If $k_\psi(\cdot)$ depends only on the magnitude of d, then so does the covariance function $c(\cdot)$, implying that $z(s)$ is isotropic. If the kernel is square-integrable, for a given $c(\cdot)$, $k_\psi(\cdot)$ can be obtained as the inverse Fourier transform of the square root of the spectral density of c (Thiébaux and Pedder, 1987). An example of a flexible family of isotropic kernels is obtained from the Matèrn class of isotropic correlations (Matérn, 1986), which is known to correspond to processes with widely different degrees of smoothness. The spectral density of the correlation function in the $\mathbb{R}^2$ Matèrn class, with range $\lambda > 0$ and smoothness $\nu > 0$, is given by $f(\omega) \propto 1/(\lambda^2 + \omega^2)^{\nu+1}$. The corresponding kernel is the inverse Fourier transform of $1/(\lambda^2 + \omega^2)^{\nu/2+1/2}$, which is proportional to

$$(\lambda s)^{\nu/2-1/2} \mathcal{K}_{\nu-1/2}(\lambda s), \quad \lambda > 0, \nu > 1, \tag{2}$$

where $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν (see Abramowitz and Stegun, 1965).

In practice Equation (1) is discretized over a regular grid. So we obtain the approximation

$$z(s) \approx \sum_u k_\psi(s - u) v_u. \tag{3}$$

Since the grid that indexes the variables v_u is usually of much smaller size than the data set, discrete process convolutions can be used to fit large dimensional problems practically. In fact computational demands are simplified by greatly reducing the parameter dimensions (Lee et al., 2002). Flexible choices of smoothing kernels can be considered for Gaussian processes (Barry and Ver Hoef, 1996; Kern, 2000; Hoef et al., 2004), they can be also extended easily to non-Gaussian cases by simply convolving non-Gaussian variables.

3 Statistical models for MOZART output

We base our representation of the stochastic process that corresponds to MOZART output on Equation (3). As discussed in the introduction, MOZART produces output for lattices of 64×128 for each one of 19 altitudes. In this application we focus on just the height that corresponds to 300mb. Furthermore, we are interested in the in the annual MOZART simulated spring ozone averages (March, April, May). Let $y_{t,s}$ be the log-transformed spring ozone average for year $t = 1, \ldots T$ and location $s = (s_x, s_y)$, where s_x denotes longitude and s_y denotes latitude. Let N denote the number of such locations. We assume that there is no annual trend and no correlation between the spring averages of different years. Thus, for each cell, we have T independent replicates of average spring ozone.

To obtain a model based on Equation (3) we define a regular lattice $(u_1, \ldots, u_U)$. The definition of the lattice affects the spatial resolution of the model. Lattices with high number of points will have higher resolution and better ability to capture small spatial variability than coarser lattices. This comes at the cost of a larger number of parameters. The application at hand is mostly about global comparisons, so high resolution is not a pressing issue. Also, given the size of the dataset a large U will make the model computationally infeasible. We use a grid of 10×20, so that $U = 200$. Each grid cell in the lattice corresponds to about 41 MOZART output cells.

We assume that v_u are normally distributed exchangeable latent variables, taking values on the lattice. Then, for each location s and time t,

$$y_{t,s} = \mu_s + \sum_u k_{\psi,\nu}(s-u)v_u + \epsilon_{t,s},\ \epsilon_{t,s} \sim N(0,\tau_s^2), \quad v_u \sim N(0,\eta^2). \tag{4}$$

We notice that there is a location and a variance parameter, μ_s and τ_s^2, for each MOZART grid cell. This mimics the exploratory data analysis based on a graphical comparison of maps of grid cell means and variances. The kernel in Equation (4) can be chosen so that the latitudinal and longitudinal ranges are different. We denote the ranges as (ψ_x, ψ_y) and let $\phi_x = \omega^2\psi_x$ and $\phi_y = \omega^2\psi_y$, where ω is the distance between two nearest lattice points. We use a distance that accounts for the curvature and the spherical shape of the earth. We consider three kernel functions, Gaussian, rational quadratic and Matérn kernels.
The Gaussian kernel is,

$$\frac{1}{2\pi\sqrt{\phi_x\phi_y}} \exp\{-\frac{\|s_x-u_x\|^2}{2\phi_x} - \frac{\|s_y-u_y\|^2}{2\phi_y}\}.$$

The rational quadratic kernel is,

$$\left(1+\frac{\|s_x-u_y\|^2}{\phi_x}\right)^{-\nu}\left(1+\frac{\|s_y-u_x\|^2}{\phi_y}\right)^{-\nu}, \nu > 0.$$

The Matérn kernel is,

$$\left(\frac{1}{\Gamma(\nu)2^{(\nu-1)}}\right)^2\left(\frac{4\nu\|s_x-u_x\|\|s_y-u_y\|}{\sqrt{\phi_x\phi_y}}\right)^{\nu} K_\nu\left(\frac{2\sqrt{\nu}\|s_x-u_x\|}{\sqrt{\phi_x}}\right) K_\nu\left(\frac{2\sqrt{\nu}\|s_y-u_y\|}{\sqrt{\phi_y}}\right), nu > 0.$$

Here

$$||s_x-u_x|| = 3,963.34 \times \theta_x \quad \text{and} \quad ||s_y-u_y|| = 3,963.34 \times |s_y-u_y|$$

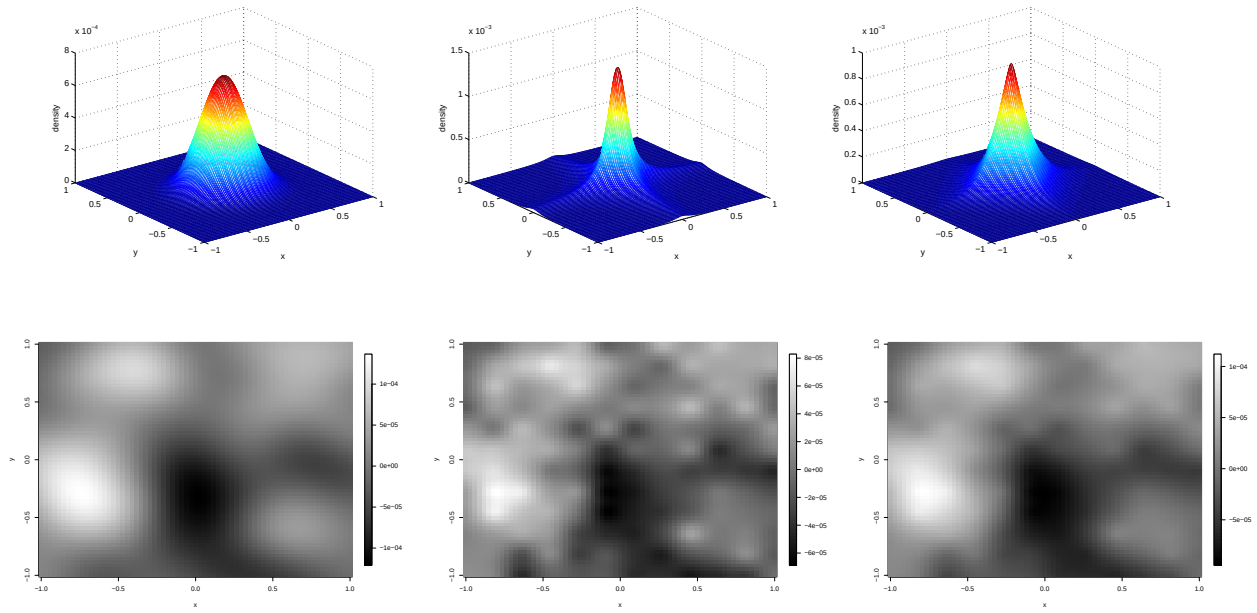
$$\text{where} \quad \theta_x = \begin{cases} |s_x-u_x| & \text{if} \quad |s_x-u_x| \in [0,\pi] \\ 2\pi - |s_x-u_x| & \text{if} \quad |s_x-u_x| \in (\pi, 2\pi] \end{cases}$$

where miles have been used as the unit for distance.

When compared to the Gaussian kernel, the rational quadratic can have much longer tails, for small values of ν. This provides the ability to capture dependencies at fairly large ranges. For the Matérn kernel, ν is responsible for the smoothness of the process. In Figure 1 the upper row shows the shapes of all three comparable kernels with similar interquartile ranges. For the Gaussian kernel, $\sigma_x^2 = \sigma_y^2 = 0.051$. For the rational quadratic kernel, $\sigma_x^2 = \sigma_y^2 = 0.01$ and $\nu = 0.75$. For the Matérn kernel, $\sigma_x^2 = \sigma_y^2 = 0.27$ and $\nu = 0.75$. The bottom row of Figure 1 shows the resulting processes by convolving a white noise process with all three kernel functions under consideration. The process convolution convolving with a Gaussian kernel produces a much smoother random field than that convolving with either a rational quadratic kernel or a Matérn kernel, for small values of ν.

To complete the model in Equation (4) we specify vague priors for τ_s^2 and η^2 in the form of inverse gamma distributions with shape and scale parameters equal to 1. Priors for the kernel parameters are specified as $\psi_x, \psi_y \sim \text{Unif}(0,2)$, $\nu \sim \text{Unif}(0,8)$.

To perform the comparison of the two sets of MOZART output we proceed by fitting the model separately to each data set. This produces two sets of posterior distributions. Denote Θ the



(a) Gaussian kernel (b) Rational quadratic kernel (c) A Matérn kernel

Figure 1: Shapes of the three considered kernels: Gaussian, rational quadratic, and Matérn (top panels), and simulations obtained by convolving a white noise process with those kernels (bottom panels).

collection of all model parameters. Denote $\boldsymbol{Y}_C$ the MOZART output obtained from CAM forcings and $\boldsymbol{Y}_N$ the one obtained from NCEP forcings. Then the predictive posterior distribution of a new set of output $\tilde{\boldsymbol{Y}}$ conditional on output $\boldsymbol{Y}_i$ is

$$p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_i) = \int_\Theta p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_i, \Theta) p(\Theta|\boldsymbol{Y}_i), \quad i = C, N.$$

If both $\boldsymbol{Y}_C$ and $\boldsymbol{Y}_N$ produce similar predictions, then we can assume that they are equivalent. This can be decided on the basis of the similarities between $p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_C)$ and $p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_N)$.

An alternative assessment of the MOZART output is done using a model choice framework. We consider two models. Model 1, denoted as $\mathcal{M}_1$, corresponds to a likelihood given by a common set of parameters for all observations, say, $p(\boldsymbol{Y}_C, \boldsymbol{Y}_N|\Theta)$. Model 2, denoted as $\mathcal{M}_2$, corresponds to a likelihood specified by two different sets of parameters, say $p(\boldsymbol{Y}_C, \boldsymbol{Y}_N|\Theta_C, \Theta_N) = p(\boldsymbol{Y}_C|\Theta_C)p(\boldsymbol{Y}_N|\Theta_N)$. The decision of whether the two MOZART output are equivalent corresponds to $\mathcal{M}_1$ being the preferred model.

3.1 Model fitting

We fit the model by running a customized Markov chain Monte Carlo (MCMC) (see, for example Gamerman and Lopes, 2006) that produces samples of the joint posterior distribution of all parameters. In order to obtain faster convergence, better mixing and maintain the simplicity of our Monte Carlo methods, the MCMC is based on grouping the parameters in blocks.

The first block consists of μ_s and τ_s^2 for each location site. For these two parameters, we write the full conditional distribution as $p(\mu_s, \tau_s^2|\ldots) = p(\tau_s^2|\mu_s, \ldots)p(\mu_s|\ldots)$, where $\ldots$ is a short hand for all other parameters and observations. We sample from such a distribution using Metropolis-Hastings steps (Gamerman and Lopes, 2006). At the i-th iteration of the MCMC, we obtain a

proposed value for μ_s, say μ_s^*, as $\mu_s^* = \mu_s^{(i-1)} + \zeta,\ \zeta \sim N(0, \xi_\mu^2)$. Here $\mu_s^{(i-1)}$ is the current value of μ_s. μ_s^* will be either accepted or rejected based on the Metropolis-Hasting ratio

$$\min\left\{1, \left(\frac{\sum_t \left(y_{t,s} - \mu_s^{(i-1)} - \sum_u k_{\psi,\nu}(s-u)v_u\right)^2 + 2\beta_\tau}{\sum_t \left(y_{t,s} - \mu_s^* - \sum_u k_{\psi,\nu}(s-u)v_u\right)^2 + 2\beta_\tau}\right)^{(\frac{T}{2}+\alpha_\tau)}\right\}.$$

If the proposed μ_s^* is accepted, generate a corresponding τ_s^{2*} from

$$Inv.Gam\left(\frac{T}{2} + \alpha_\tau, \sum_t \frac{\left(y_{t,s} - \mu_s^* - \sum_u k_{\psi,\nu}(s-u)v_u\right)^2}{2} + \beta_\tau\right).$$

Otherwise leave both μ_s and τ_s^2 unchanged.

Once μ_2 and τ_s^2 have been updated, for all s, we consider the parameters ψ_x, ψ_y, ν and v_u, for all u. Let $V = (v_1, \ldots v_U)'$. Then we write the joint full conditional as $p(\psi_s, \psi_y, \nu, V|\ldots) = p(V|\psi_x, \psi_y, \nu, \ldots)p(\psi_x, \psi_y, \nu|\ldots)$. We obtain candidate values, ψ_x^*, ψ_y^*, and ν^* for ψ_x, ψ_y, and ν respectively, using a random walk, in a manner analogous to what was done for μ_s. Let

$$\hat{V} = (H'_{\hat{V}} M_{\hat{V}}^{-1} H_{\hat{V}})^{-1}(H'_{\hat{V}} M_{\hat{V}}^{-1} W_{\hat{V}}), \quad W_{\hat{V}} = \begin{pmatrix} \sum_t y_{t,S}/T - \mu_S \\ \mathbf{0}_U \end{pmatrix},$$

where $\sum_t y_{t,S}/T - \mu_S$ denotes the vector of dimension equal to the number of MOZART output grid cells and components $\sum_t y_{t,s}/T - \mu_s$. $\mathbf{0}_U$ denotes the vector of U components equal to 0.

$$H_{\hat{V}} = \begin{pmatrix} k_{\psi,\nu}(s_1 - u_1) & \ldots & k_{\psi,\nu}(s_1 - u_U) \\ \vdots & \ddots & \vdots \\ k_{\psi,\nu}(s_S - u_1) & \ldots & k_{\psi,\nu}(s_S - u_U) \\ 1 & \ldots & 0 \\ & \ddots & \\ 0 & \ldots & 1 \end{pmatrix}, \quad \text{and} \quad M_{\hat{V}} = \begin{pmatrix} \tau^2 I_S & 0 \\ 0 & \eta^2 I_U \end{pmatrix},$$

where I_a denotes the $a \times a$ identity matrix. ψ_x^*, ψ_y^*, and ν^* will be accepted based on the Metropolis-Hasting ratio

$$\min\left\{1, \exp\left(-\frac{1}{2\eta^2}\sum_u (\hat{v}_u^{*2} - \hat{v}_u^{(i-1)2})\right) \times \exp\left(-\frac{1}{2}\sum_{t,s} \frac{(y_{t,s} - \mu_s - \sum_u k_{\psi^*,\nu^*}(s-u)\hat{v}_u^*)^2}{\tau_s^2}\right) \times \right.$$
$$\left. \times \exp\left(\frac{1}{2}\sum_{t,s} \frac{(y_{t,s} - \mu_s - \sum_u k_{\psi^{(i-1)},\nu^{(i-1)}}(s-u)\hat{v}_u^{(i-1)})^2}{\tau_s^2}\right)\right\}.$$

If the proposed ψ_x^*, ψ_y^*, and ν^* are accepted, V is updated by sampling from a U-variate normal with mean $\hat{V}$ and covariance matrix $(H'_{\hat{V}} M_{\hat{V}}^{-1} H_{\hat{V}})^{-1}$. Since we can tune the proposal distribution of ψ_x, ψ_y and ν to obtain an acceptance rate between 40% and 60%, the computations involving large matrices required to update V are performed on a fraction of the total number of iterations. This is an additional important advantage of the blocking scheme that we use.

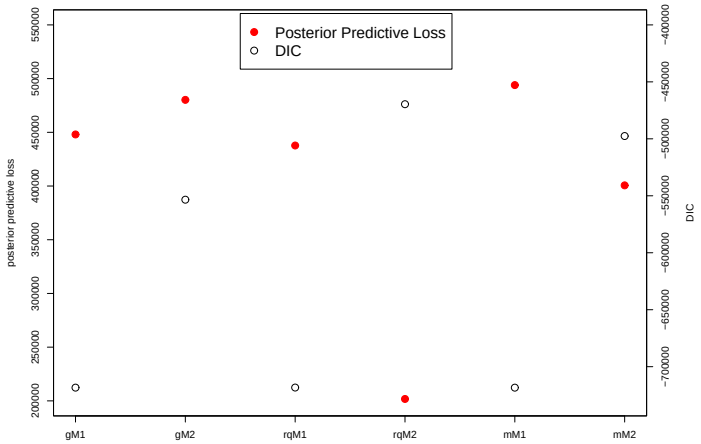


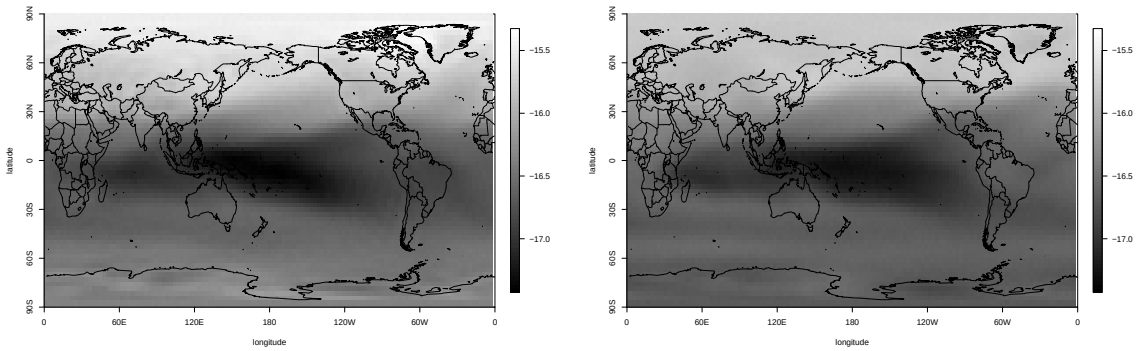
Figure 2: Results of posterior predictive loss and DIC values when convolving with all three kernels for $\mathcal{M}_1$ and $\mathcal{M}_2$. (g stands for Gaussian, rq stands for rational quadratic, m stands for Matérn).

4 Results

Conceptually, one of the most appealing ways of performing Bayesian model comparisons is to use Bayes factors (Kass and Raftery, 1995). Unfortunately in this application we have little prior knowledge on some of the parameters, we resort to using vague priors. Thus Bayes factors are not appropriate for our application. Additionally, the model is already very computationally demanding and the calculation of the marginal distributions needed for the Bayes factors increases the amount of computations substantially. Thus we focus on criteria for model comparison problems that can be directly calculated from the MCMC output and have little sensitivity to the prior used. The methods used for the comparison of models $\mathcal{M}_1$ and $\mathcal{M}_2$ are: posterior predictive loss (Gelfand and Ghosh, 1998) and Deviance Information Criterion (Spiegelhalter et al., 2002). The details of the criteria used for comparison are given in the following sections. To study the similarities between $p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_C)$ and $p(\tilde{\boldsymbol{Y}}|\boldsymbol{Y}_N)$ we calculate the percentage of overlapping of the posterior intervals corresponding to each grid cell and Bhattacharyya distance (Bhattacharyya, 1943).

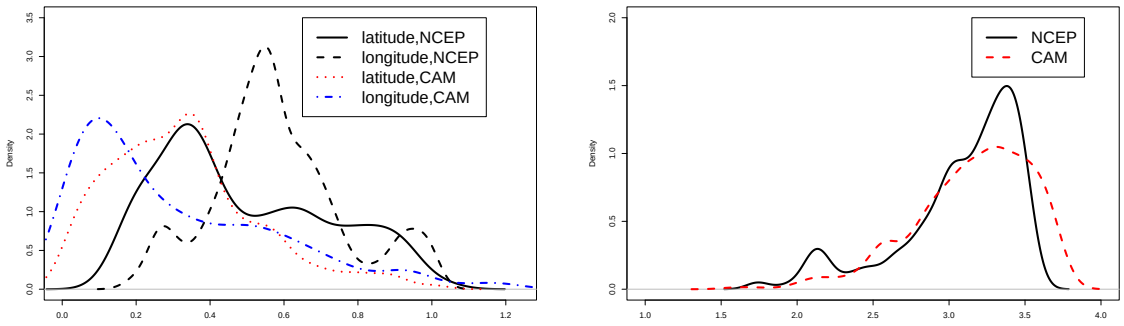
We start our analysis by considering the problem of choosing one of the three proposed kernel families. The results presented in this paper are based on 1,000 iterations after a burn-in of 14,000. We fit models $\mathcal{M}_1$ and $\mathcal{M}_2$ using each one of the kernels and compared the six resulting models using DIC and posterior predictive loss. The results are shown in Figure 2. We observe that models based in the rational quadratic kernel minimize the criteria. So for the rest of the paper we focus on that kernel.

Figure 3 shows the posterior predictive means of $\tilde{y}_s$ for NCEP and CAM. This graphical comparison shows only subtle differences, mainly in the areas that correspond to the poles. For a better understanding of the differences between the two outputs we need to compare features of the posterior predictive distributions other than the mean. For a comparison of the kernel parameter values we have in Figure 4 the posterior distributions of ψ_x, ψ_y, and ν. We notice that there are some latitudinal-longitudinal differences for the scaling parameters ψ. However the smoothing parameter ν has a similar posterior distribution for NCEP and CAM. To assess the goodness of fit, we calculated the errors based on the posterior predictive means. Figure 5 shows the maps of such errors as well as the corresponding histograms. We observe that there is no clustering and no patches and the histograms are symmetric and short tailed.



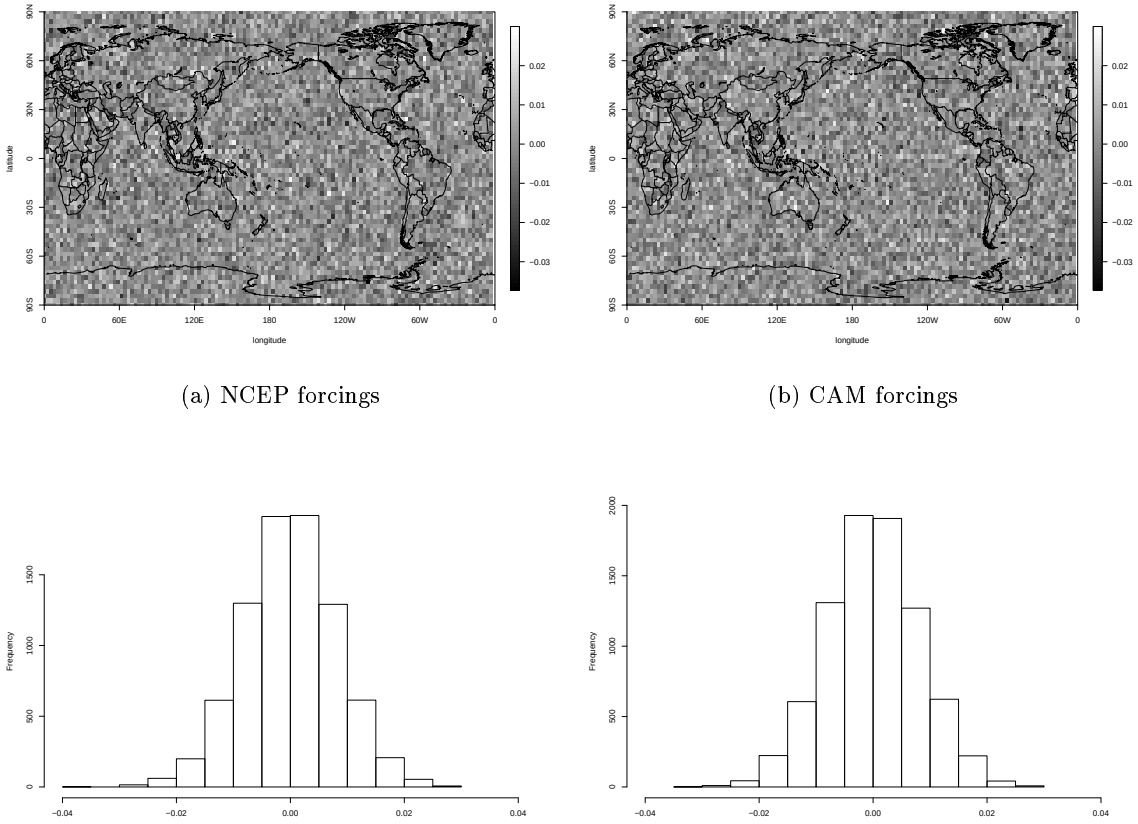
(a) NCEP forcings (b) CAM forcings

Figure 3: Posterior predictive means of $\tilde{y}_s$ based on the model that uses a rational quadratic kernel.



(a) Posterior distributions of ψ's (b) Posterior distribution of ν's

Figure 4: Posterior distributions of the parameters of the rational quadratic kernels.



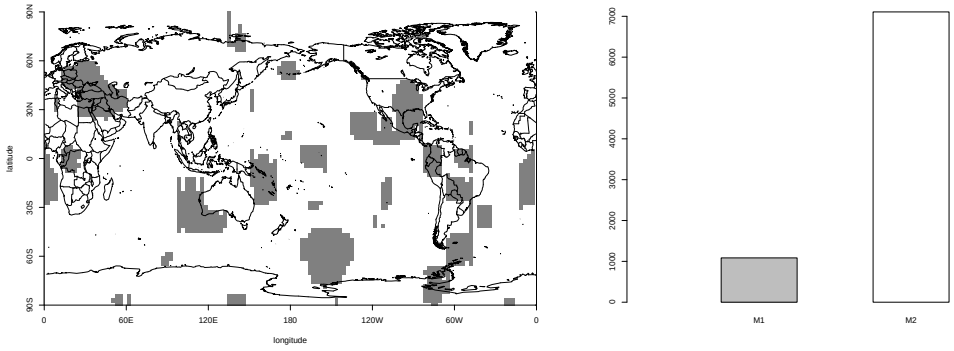
(a) NCEP forcings

(b) CAM forcings

(c) NCEP forcings

(d) CAM forcings

Figure 5: Predictive errors, spatial distributions and histograms.



(a) Image plot of the selected model

(b) Counts of cells where a given model is selected

Figure 6: Model selection based on minimizing posterior predictive loss. (gray color stands for $\mathcal{M}_1$, white color stands for $\mathcal{M}_2$).

4.1 Posterior Predictive Loss

The posterior predictive loss criterion was originally proposed in Gelfand and Ghosh (1998). The criterion chooses the model that minimizes the loss incurred in estimating the observations using the posterior predictive distribution. We use a quadratic loss which results in

$$\begin{aligned} L_s(\mathcal{M}_k) &= \sum_{t=1}^{T} E_{y_{rep}|y_{obs}}(y_{s,t,rep} - y_{s,obs})^2 \\ &= \sum_{t=1}^{T} \{v_s(y_{rep}|y_{obs}) + (\mu_s(y_{rep}|y_{obs}) - y_{s,t,obs})^2\}, \end{aligned}$$

where,

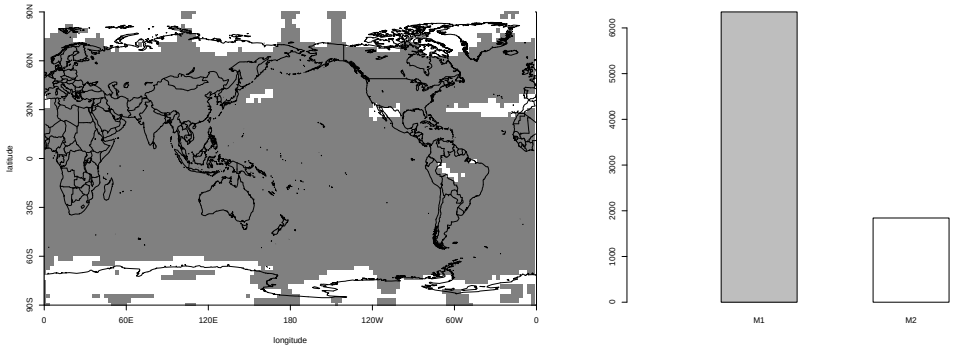
$$v_s(y_{rep}|y_{obs}) = \frac{1}{M}\sum_{i=1}^{M}(\tau_s^{2(i)} + K_s'^{(i)}K_s^{(i)}\eta^{2(i)}), \quad \mu_s(y_{rep}|y_{obs}) = \frac{1}{M}\sum_{i=1}^{M}\mu_s^{(i)},$$

where the super index (i) denotes the i-th iteration of the MCMC, obtained under $\mathcal{M}_k$ and M is the total number of available samples. Notice that we calculate one value of the criterion per grid cell. This permits the identification of the sites where the model output differs for different forcings. The overall criterion is given by the sum of all L_s.

Subfigure 6(a) shows a global map of selected models (gray color stands $\mathcal{M}_1$, white color stands for $\mathcal{M}_2$) which minimize the posterior predictive losses. $\mathcal{M}_2$ is the model that minimizes the posterior predictive loss at most locations. Subfigure 6(b) provides an idea of the proportion of grid cells where the criteria is minimized by each model.

4.2 Deviance Information Criterion

Two popular criteria for model comparison are the Akaike Information Criterion (AIC) (Akaike, 1973, 1974) and the Bayesian Information Criterion (BIC)) (Schwarz, 1978). They both require the specification of the number of parameters in the model as well as the calculation of the maximum of the posterior distribution. In our case, it is not clear how to count the effective number of parameters for process convolutions, e.g., for the latent process, the count is probably somewhere



(a) Image plot of the selected model

(b) Counts of cells where a given model is selected

Figure 7: Results for the rational quadratic kernel: minimizing DIC values. (gray color stands for $\mathcal{M}_1$ and white color stands for $\mathcal{M}_2$)

between 1 and U because of spatial dependence among those U parameters. Deviance Information Criterion (DIC) is defined in Spiegelhalter et al. (2002) as an attempt to produce a criterion that incorporates an implicit definition of the effective number of parameters. The popularity of DIC is also due to the fact that it can be directly calculated from MCMC output with very few additional calculations. For insight on the pros and cons of the DIC we refer the reader to the discussion in Spiegelhalter et al. (2002). One of the strongest criticism being that of using the data twice.

DIC considers both the goodness of fit and the model complexity. For our application we have

$$DIC_s(\mathcal{M}_k) = 2\overline{D_s(\theta)}_k - D_s(\bar{\theta})_k, \quad k = 1, 2$$

where, dropping the dependence on the model index k for clarity,

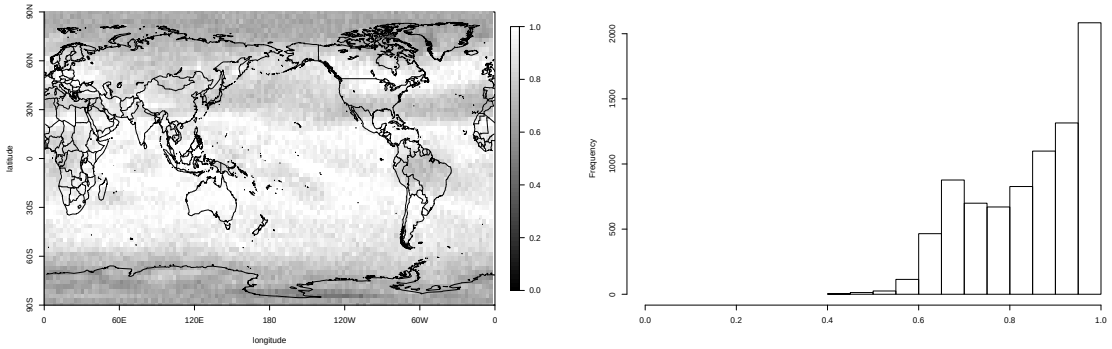
$$D_s(\theta) = T\log(2\pi\tau_s^2) + 2\sum_{t=1}^{T} \frac{(y_{t,s} - \mu_s - \sum_u k(s-u)v_u)^2}{2\tau_s^2}.$$

So, in analogy to the posterior predictive loss criterion, we have a comparison for each grid cell. The overall criterion is given as the sum of all grid cell values. Also, as for the previous criterion, the preferred model is the one that minimizes the DIC.

Subfigure 7(b) shows results that are in strong contradiction to those in Subfigure 6(b). It indicates that $\mathcal{M}_1$ is preferred to $\mathcal{M}_2$ at a large number of locations. We prefer the posterior predictive loss approach since DIC fails for latent variable models (Celeux et al., 2006). Moreover, since we obtain some negative values for the effective number of parameters, the DIC approach is not trustworthy. We observe that the images obtained from both the posterior loss approach and the DIC approach are very patchy that is because of the qualitative property of both approaches.

4.3 Overlapping of probability intervals

The Bayesian approach provides uncertainty measurements on parameter estimates in the form of distribution. This is also true for the posterior predictive distributions, $p(\tilde{y}_s|\boldsymbol{Y}_k), k = C, N$, for each grid cell. From MCMC runs, we can obtain the 95% credible intervals that correspond to the quantiles 0.025 and 0.975 of the distribution of $\tilde{y}_s$, conditional on either $\boldsymbol{Y}_C$ or $\boldsymbol{Y}_N$. We can



(a) (b)

Figure 8: Percentage of overlaps between 95% posterior predictive probability intervals obtained conditioning on either $\boldsymbol{Y}_C$ or $\boldsymbol{Y}_N$.(a) Percentages for each cell. (b) Histogram of the percentages calculated in (a).

quantify the agreement between these two distributions by calculating the percentage of overlap of such intervals. We repeat the calculation for each s to obtain a map of the overlaps. This is shown in Figure 8. We observe that the overlap is high between the tropics, with the exception of Brazil. It somewhat low in the Northern Pacific and Atlantic and it is fairly low in the poles. An overall measure of overlap is given by the histogram in right panel of Figure 8. This indicates that for most of the cells there is at least an overlap of 60% in the 95% posterior predictive probability intervals.

4.4 Bhattacharyya Distance

An alternative way of quantifying the agreement between the two predictive distributions available at each grid cell is to consider some measure of discrepancy between distributions. Some possibilities are given by Bhattacharyya distance (Bhattacharyya, 1943), Hellinger distance (Matusita, 1955; Rao, 1963; Beran, 1977) and Kullback-Leibler distance (Cover and Thomas, 1991). We consider all of these and found that the first one is the most numerically stable of the three.

The Bhattacharyya distance is defined by,

$$B(f_1|f_2) = \int \sqrt{f_1(x)f_2(x)}dx.$$

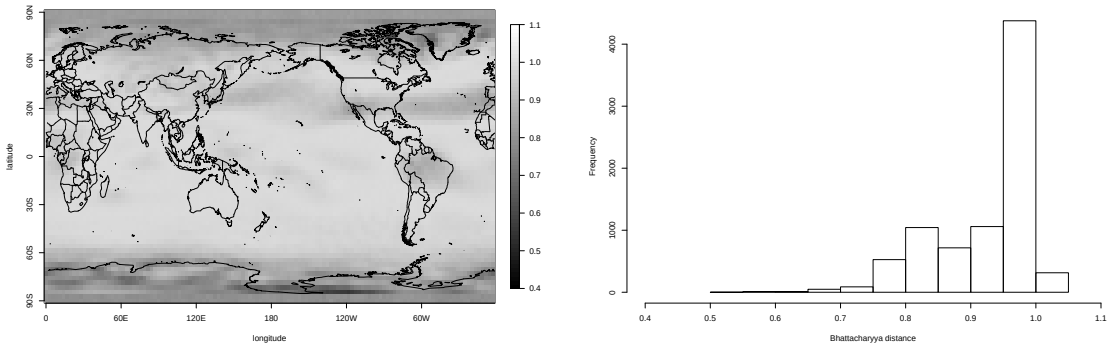
It has a one to one relationship with the Hellinger distance, which is defined as

$$H(f_1|f_2) = \int (\sqrt{f_1(x)} - \sqrt{f_2(x)})^2 dx = 2 - 2\int \sqrt{f_1(x)f_2(x)}dx = 2(1 - B(f_1|f_2)).$$

Thus the larger the Bhattacharyya distance is, the more similar the two distributions are.

From the definition of Bhattacharyya distance we observe that

$$B(f_1|f_2) = \int \sqrt{\frac{f_2(x)}{f_1(x)}} f_1(x)\, dx \approx \frac{1}{M}\sum_{i=1}^{m} \sqrt{\frac{f_2(x_1^{(i)})}{f_1(x_1^{(i)})}}, \quad x_1^{(i)} \sim F_1, \tag{5}$$



(a) (b)

Figure 9: $B(p(\tilde{y}_s|\boldsymbol{Y}_C)|p(\tilde{y}_s|\boldsymbol{Y}_N))$. (a) Estimated values for each cell. (b) Histogram of the percentages calculated in (a).

so $B(f_1|f_2)$ can be effectively calculated using a Monte Carlo approach. Analogous calculations can be done for the Hellinger and the Kullback-Leibler distances, but they will not result in expectations of positive quantities. Thus, the positiveness of the estimation is not guaranteed. We encountered this problem in our application.

To assess the similarity of NCEP and CAM based simulations we calculated the Bhattacharyya distance for the posterior predictive densities $p(\tilde{y}_s|\boldsymbol{Y}_C)$ and $p(\tilde{y}_s|\boldsymbol{Y}_N)$, using MCMC output. In Equation 5 we can exchange f_1 with f_2. In theory this produces the same value, but in practice we obtain two different approximations to $B(f_1|f_2)$. To avoid this problem we average the two estimates. The results are presented in Figure 9. We observe that the overall features agree with those in Figure 8 but that the field measuring the agreement between the predictives is smoother in this case.

5 Conclusions

We have presented a method for the comparison of large data sets obtained from computer simulators different sets of inputs. The use of process convolutions facilitates the study of the random fields by reducing the dimensionality of the problem. The use of a Bayesian approach allows for a quantification of the uncertainties in terms of probability distributions. These provides a precise assessment of the similarities between predictions that incorporates all the parameter estimation variabilities.

Considering the problem from a model comparison setting is of little use in this application. By applying a 0-1 rule we obtain very patchy results that do not correspond with the evidence provided by the model predictions. A more useful analysis is obtained by considering some measures of discrepancy between predictive distributions. We proposed a measure that is based on the overlap of predictive intervals. We believe that this is an intuitive measure of discrepancy that has the advantage of corresponding to percentages. Similar comparisons could have been calculated for the posterior distributions of μ_s or τ_s^2, or their joint posterior distribution. Since the predictive posterior indicates how well the model is able to reproduce the data that were actually observed, we believe that it provides a better summary of the fit.

The present application was developed focusing on a specific altitude. A comprehensive assess-

ment of MOZART output requires consideration of whole profiles for each grid cell. Additionally, we have considered only one season and assumed independence between observations corresponding to different years. A full spatio-temporal model that considered the sequential structure of the simulations would also be a desirable extension of the present model.

6 Acknowledgments

We want to acknowledge the fruitful collaboration with Dr Hess and Dr Mahowald of NCAR who proposed the problem and provided the data. The authors were partially supported by National Science Foundation grants ATM 0405451 and DMS 0504851.

References

Abramowitz, M. and I. Stegun. 1965. *Handbook of Methematical Functions.* New York: Dover Publications, Inc.

Akaike, H. 1973. Information Theory and an Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*, eds. B. N. Petrov and F. Cáski, 267–281. Budapest: Akademiai Kaidó.

—. 1974. A New Look at Statistical Model Identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Barry, R. P. and J. M. Ver Hoef. 1996. Blackbox Kriging: Spatial Prediction Without Specifying Variogram Models. *Journal of Agricultural, Biological, and Environmental Statistics* 1: 297–322.

Beran, R. 1977. Minimum Hellinger Distance Estimates for Parametric Models. *Ann. Statist.* 5(3): 445–463.

Bhattacharyya, A. 1943. On a Measure of Divergence between Two Statistical Populations Defined by Their Distributions. *Bull. Calcutta Math. Soc.* 35: 99–110.

Celeux, G., F. Forbes, C. Robert, and D. Titterington. 2006. Deviance Information Criteria for Missing Data Models. *Bayesian Analysis* 1(4): 00–00.

Cover, T. M. and J. A. Thomas. 1991. *Elements of Information Theory.* New York: Wiley.

Gamerman, D. and H. Lopes. 2006. *Markov Chain Monte Carlo.* 2nd ed. London, UK: Chapman and Hall.

Gelfand, A. E. and S. K. Ghosh. 1998. Model Choice: A Minimum Posterior Predictive Loss Approach. *Biometrika* 85(1): 1–11.

Higdon, D. 2002. Space and Space-time Modeling Using Process Convolutions. In *Quantitative Methods for Current Environmental Issues*, eds. C. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37–56. London: Springer-Verlag.

Hoef, J. V., N. Cressie, and R. Barry. 2004. Flexible Spatial Models Based on the Fast Fourier Transform (FFT) for Cokriging. *Journal of Computational and Graphical Statistics* 13: 265–282.

Kalnay, E., M. Kanamitsu, R. Kistler, W. Collins, D. Deaven, L. Gandin, M. Iredell, S. Saha, G. White, J. Woollen, Y. Zhu, A. Leetmaa, and B. Reynolds. 1996. The NCEP/NCAR 40-Year Reanalysis Project. *Bulletin of the American Meteorological Society* 77(3): 437–471.

Kasagara, A. and W. M. Washington. 1967. NCAR Global General Circulation Model of the Atmosphere. *Monthly Weather Review* 95(7): 389–402.

Kass, R. and A. Raftery. 1995. Bayes Factors and Model Uncertainty. *Journal of the American Statistical Association* 90: 773–795.

Kern, J. C. 2000. *Bayesian Process-Convolution Approaches to Specifying Spatial Dependence Structure.* Ph.D. Thesis. Duke University, Durham, NC 27708.

Lee, H., D. Higdon, Z. Bi, M. Ferreira, and M. West. 2002. Markov Random Field Models for High-Dimensional Parameters in Simulations of Fluid Flow in Porous Media. *Technometrics* 44: 230–241.

Matérn, B. 1986. *Spatial Variation (2nd ed; Lecture Notes in Statistics Vol. 36).* New York: Springer-Verlag.

Matusita, K. 1955. Decision rules on the distance, for problems of fit, two-samples, and estimation. *Ann. Math. Statist.* 26: 631–640.

Rao, C. R. 1963. Criteria of estimation in large samples. *Sankhyā* 25: 189–206.

Schwarz, G. 1978. Estimating the Dimension of a Model. *Annals of Statistics 6* 461–464.

Spiegelhalter, D. J., N. G. Best, B. P. Carlin, and A. van der Linde. 2002. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society Series B* 64: 583–639.

Thiébaux, H. J. and M. A. Pedder. 1987. *Spatial Objective Analysis with Applications in Atmospheric Science.* London: Academic Press.